CURMUDGEON CORNER

Deepfakes, Simone Weil, and the concept of reading

Steven R. Kraaijeveld[1]

"And, without believing in them, everyone smiles at appearances and pretends to accept them."

—Albert Camus, Notebooks: 1935-1942

The specter of the deepfake looms over our digital worlds. The technologies that allow for the creation and dissemination of highly realistic AI-generated audiovisual images are becoming increasingly sophisticated and pervasive. People can appear to say and do things that they never said or did; phantastic scenes and events can be depicted as if they were real and actually occurred.

The possibilities of deepfakes—unreal yet highly convincing and compelling audiovisual images—are fathomless. We may be fooled, and fool each other, in countless ways. Presently, the use of deepfakes is limited more by the bounds of human creativity and will than by the affordances of the relevant technologies. This means that we must think deeply about the meaning of deepfakes.

Deepfakes raise fundamental philosophical questions about the nature of—and the relations between—representation and reality. Discussions about what is real and what is mere appearance go back at least as far as Plato and his illustrious allegory of the cave. I do not want to focus on Plato, however, but rather on someone who was inspired by Plato much later; namely, the French philosopher, mystic, and activist Simone Weil (1909–1943).

In order to shed light on the shadowy status of deepfakes as they seep into our digital worlds, I want to draw particular attention to one of Weil's essays, written in the spring of 1941 and not published until 1946, called The Concept of Reading (*Essai sur la notion de lecture*). In this essay, Weil attempts to define "a concept that has not yet found a suitable name," but which she suggests is best referred to as *reading* (2015, 21).

✉ Steven R. Kraaijeveld
  s.r.kraaijeveld@amsterdamumc.nl

1 Department of Ethics, Law & Medical Humanities, Amsterdam UMC, Amsterdam, The Netherlands

I contend that we should keep Weil's concept of reading in mind when we think about and engage with deepfakes—and, more generally, when we find ourselves faced with digital environments where audiovisual illusions and confusions are increasingly entrenched. Given the rapid advances in artificial reality (AR) and virtual reality (VR) technologies, this may apply to most, if not all, digital environments in the future.

What does Weil mean by the concept of *reading*? To understand this, we must take a step back. Weil's first argument is that "sensation is immediate, a brute fact, and […] seizes us by surprise," like when a man suddenly receives a punch to the stomach, or when we touch a still-hot stove and retract our hand before knowing that our flesh is burned: "Something seizes [us] here—it is the universe, and [we] recognize it by the way it treats [us]" (2015, 21). We readily submit to the myriad sensations that overcome us.

These sensory effects on us are fairly straightforward. But there are mysteries; for instance, when seemingly insignificant sensations affect us profoundly. This is where Weil's concept of reading comes into play. "There are some black marks on a sheet of white paper; they couldn't differ more from a punch to the stomach. Yet, they can have the same effect" (2015, 22). Apparently meaningless scribbles can make us fear, suffer, rejoice.

To bring out this point more fully, Weil offers a poignant example of receiving bad news. She asks us to imagine two women, each of whom is handed a letter announcing the death of their respective son. The first woman is utterly devastated. The second woman is entirely unmoved. What could explain this difference? Did the second woman not love her son? No: the second woman cannot read.

"Everything happens to us," Weil explains, "as if the pain were in the letter itself" (2015, 22). When we read, we take the reality of the conveyed message as a given. Importantly, Weil concludes that at every instant of our lives we are "gripped from the outside, as it were, by meanings that we ourselves read in appearances" (2015, 22).

We are constantly reading the world, embodying its meanings, understanding and misunderstanding. "The sky, the sea, the sun, the stars, human beings, everything that

surrounds us is in the same way something that we read" (2015, 23). Appearance and interpretation are united for Weil in the act of reading.

This raises a crucial question. Must we always accept everything as it appears to us? What, if anything, can we do to affect our reading and to change appearances? According to Weil, we are not defenseless here. We have a set of powers that can affect appearances—even if these powers are limited, indirect, and labor-intensive.

On the one hand, Weil argues that we have a certain power over the universe. "I put a sheet of white paper over a black book and I no longer see black" (2015, 26). We have changed the view, so to speak. This power is clearly limited by our physical abilities and limitations; we can only change so much, appearance-wise. Making these kinds of alterations requires will and exertion.

On the other hand, Weil contends that we also have a certain power to change the meanings that we read in the appearances that are imposed on us. Experience and attention, for instance, can make us see things in different ways. She offers the example of an experienced captain and a passenger sailing together through a tempest. "Where the passenger reads chaos and unlimited danger, the captain reads necessities, limited dangers, resources for escaping, and an obligation to be courageous and honorable" (2015, 26). We can practice ways of looking at the world, so that we may come to read appearances differently. Weil also leaves another importance space for human beings to affect appearances. "If I see a book bound in black," she writes, "I do not doubt the black is there"—"except to *philosophize*" (2015, 24).[1] We can reflect, we can step back, we can doubt, we can philosophize. Through the exercise of our intellect, then, we may also see through appearances.

What does all of this mean for deepfakes; those compelling videos that *seem* and *feel* so real? How can these ideas help us think about and navigate digital environments, artificially generated and manipulated worlds?

First, we could try to *alter the digital world itself*, meaning that we could attempt to change or remove the very existence of deepfakes in their digital materiality. This might involve banning the creation and/or use of deepfakes, or allowing their presence only within certain parameters. We could opt for strict regulation of deepfakes, so that only some forms or applications might appear in our digital environments. Of course, while regulation may and arguably should avert some of the more egregious cases of digital deception (e.g., politically subversive or personally destructive forms), it is unlikely that digital environments will ever be free of audiovisual illusions and confusions. In fact, if one counts AR and VR technologies as fundamentally lacking

verisimilitude in ways similar to deepfakes (even if AR and VR may differ in the nature of their being outright deceptive), then tensions between appearances and reality are only likely to become more acute.

Second, we could try to *change the meaning of the appearances* of deepfakes as we encounter them. This is clearly the more difficult challenge; it requires psychological and emotional effort, perhaps more than most people are willing to exert. As Weil puts it: "The general's art is to lead enemy soldiers into reading flight in appearances and in such a way that the idea of holding fast loses all substance, all effectiveness" (2015, 26). Similarly, we need to practice reading unreality in deepfakes, individually and collectively, in such a way that the idea of responding to such audiovisual images loses all substance and effectiveness. We may be initially deceived in our reading, but let us not be further misguided into acting upon that deception.

Granted, few of us may care whether the book in front of us is ultimately the color that it appears to us. Nevertheless, when we engage with potentially deceptive audiovisual images in digital environments, we should all care whether, say, Politician X truly *said that*, or whether Celebrity Z verily *did this*. We must not read the world uncritically. After all, as Weil urges, we are not stuck in our reading of the world as if through some conditioned reflex. This goes for sun-streaked meadows and hills, rowdy streets and crowds, as much as for our digital environments with all of the gritty possibilities of deepfakes and AI-generated content. We may encounter a piece of AI-generated art and be moved by it, thinking that it expresses sincere human emotion—only to reflect that, given its artificial and insensible source, it has no meaning beyond our response to it (cf. Kraaijeveld 2024). As Weil emphasizes, we do not have to take everything before us as real. Deepfakes, digital semblances, fata morganas: we must carefully read the world.

Through attention, we must practice the art of reading in Simone Weil's particular sense. To the extent that our opinions and actions are incorrectly shaped and spurred by manipulated videos and images, we owe it to the depicted—and ultimately to ourselves—to do the work of philosophizing and to adjust our readings accordingly. We must become and remain our own ship's captains.

A correction of a sensory illusion is a modified reading, writes Weil. We must bring our reading as close to the truth as we can.

Curmudgeon Corner Curmudgeon Corner is a short opinionated column on trends in technology, arts, science and society, commenting on issues of concern to the research community and wider society. Whilst the drive for super-human intelligence promotes potential benefits to wider society, it also raises deep concerns of existential risk, thereby highlighting the need for an ongoing conversation between technology

---

[1] Emphasis mine.

and society. At the core of Curmudgeon concern is the question: What is it to be human in the age of the AI machine? -Editor.

Declarations

Conflict of interest The author declares no conflict of interest.

References

Kraaijeveld SR (2024) AI-generated art and fiction: signifying everything, meaning nothing? AI & Soc. https://doi.org/10.1007/s00146-023-01829-4

Weil S. (2015) Essay on the Concept of Reading. In *Late Philosophical Writings*, ed. Springsted EO. Translated by Springsted EO, Schmidt E. University of Notre Dame Press, Indiana: 21–27.